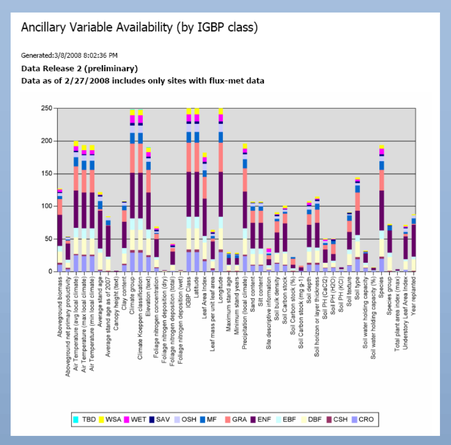
AN EVOLVING LA THUILE FLUXNET DATASET AND SUPPORT INFRASTRUCTURE

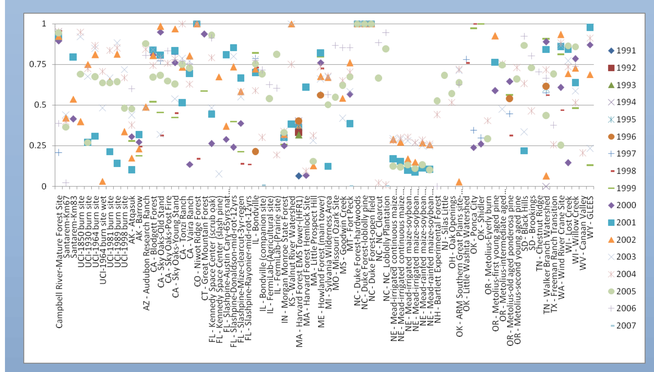
Deb Agarwal[abe] (daagarwal@lbl.gov), Dennis Baldocchi[ae] (baldocchi@nature.berkeley.edu), Monte Goode[abe] (MMGoode@lbl.gov), Marty Humphrey[j] (humphrey@cs.virginia.edu), Catharine van Ingen[d] (vaningen@microsoft.com), Dario Papale[l] (darpap@labecomail.agraria.unitus.it), Markus Reichstein[c] (mreichstein@bgc-jena.mpg.de), Matthew Rodriguez[bf] (mar008@cs.ucsd.edu), Youngryel Ryu[ae] (yryu@nature.berkeley.edu), and Rodrigo Vargas[e] (rvargas@nature.berkeley.edu)

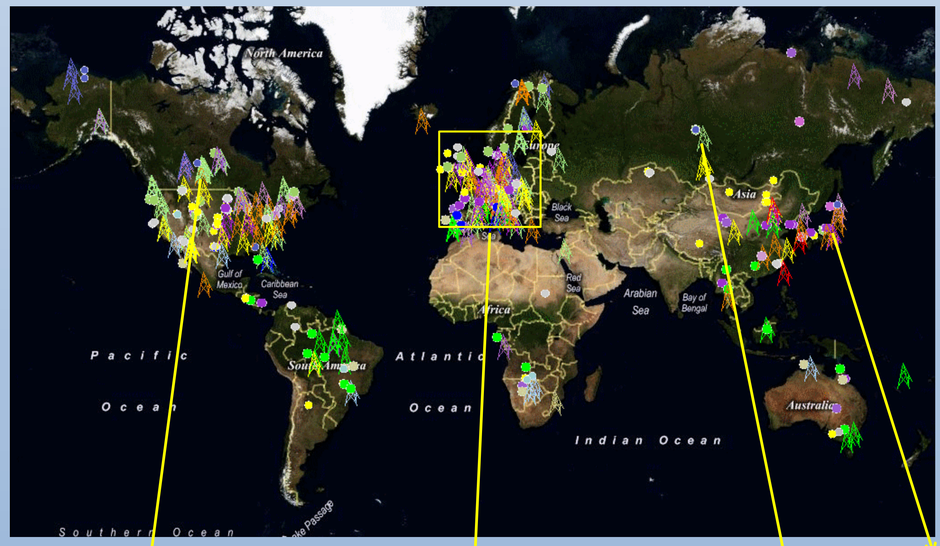
Berkeley Water Center[a]; Lawrence Berkeley National Laboratory[b]; Max Planck Institute for Biogeochemistry, Germany[c]; Microsoft Research[d]; University of California, Berkeley[e]; University of California, San Diego[f]; University of Tuscia, Italy[i]; University of Virginia[j];

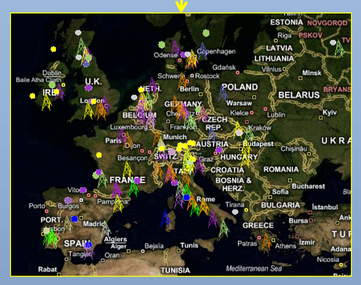
ABSTRACT

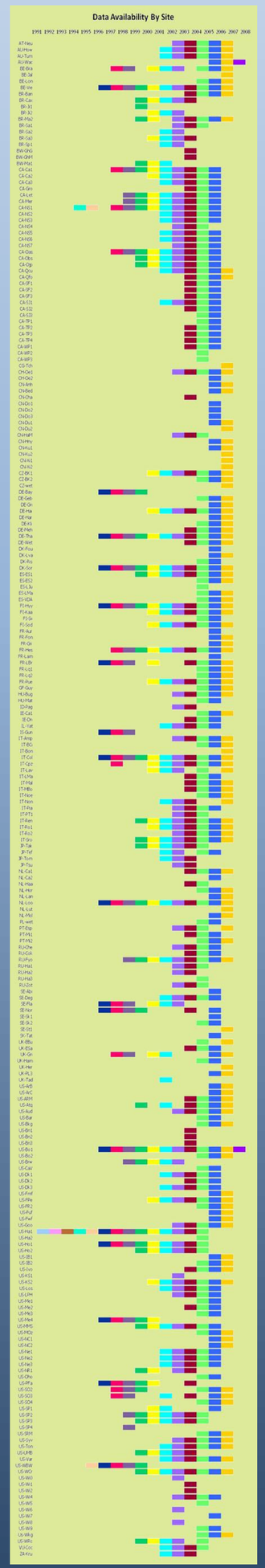
The global FLUXNET synthesis dataset contains over 960 site-years from over 250 sites and continues to grow. This growth reflects previously unavailable processed flux-met data becoming available. The ancillary site information such as biological and disturbance data also continues to evolve. As analysis efforts have begun, issues with the underlying data are often uncovered and the data are corrected when possible. The size of this dataset (90GB) makes finding site years of interest and tracking versions difficult for individual users.

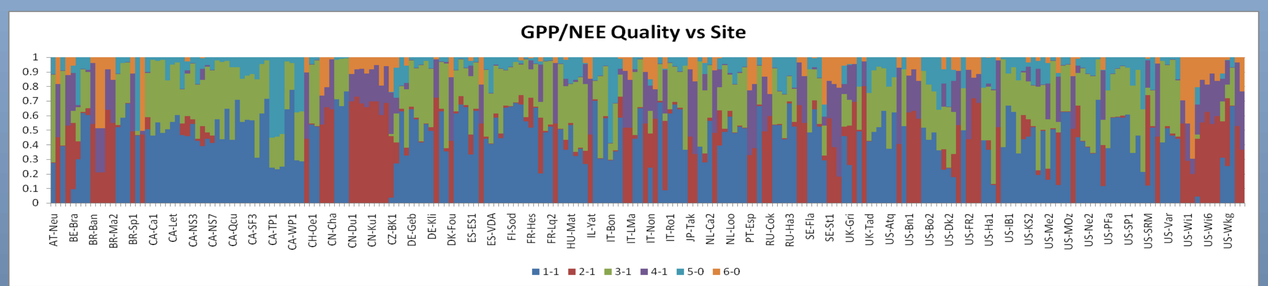


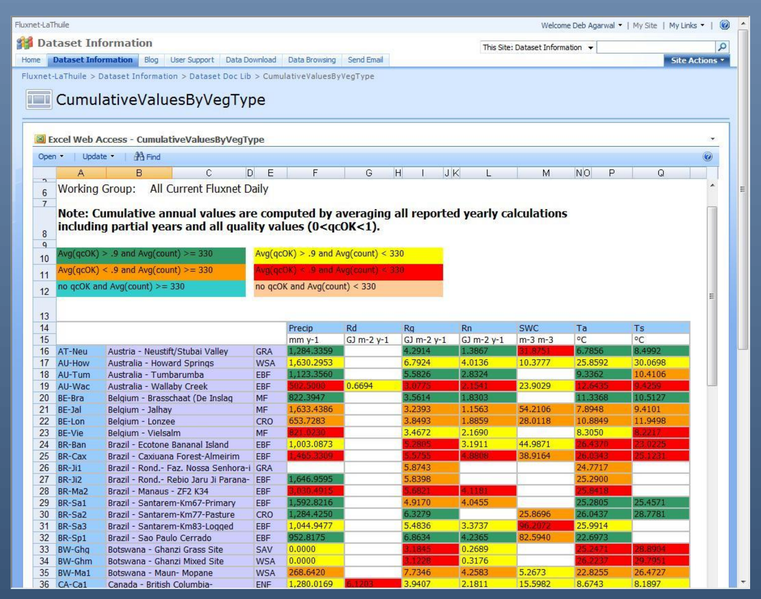


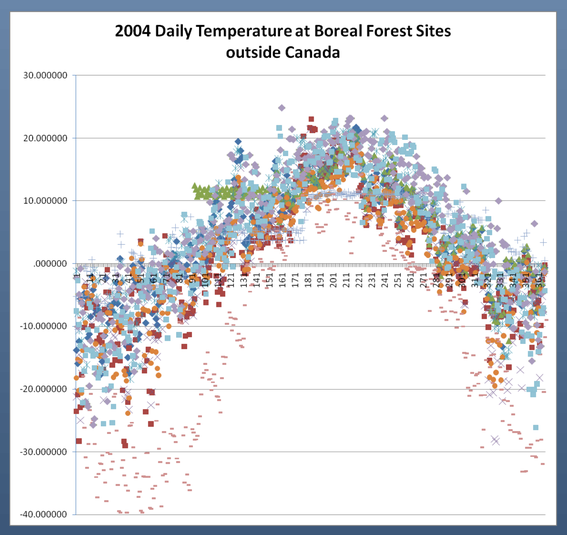












We have developed a Scientific Data Server which enables browsing of the data online, data download, version tracking, and the simple production of data summary products. The Scientific Data Server allows individual researchers to concentrate on science rather than data management. We leverage database tools such as data cubes and web reports to enable Excel pivot table, MatLab, and browser access to the data. The data cubes provide organization and aggregation of data along dimensions including time such that it is easy to retrieve daily, monthly, and yearly calculated values. In addition, we have leveraged available collaboration technology (SharePoint) to provide public pages describing the dataset, provide proposer and site PI data access, track data versions and updates, notify users of server updates, and enable support for contact between site PIs and researchers using the data.

http://www.fluxdata.org/

For further information e-mail: fluxdata-support@george.lbl.gov

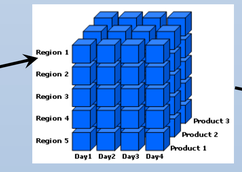
Challenge is to Connect Data, Resources, and People

Distributed Data Sets

Catalog Database

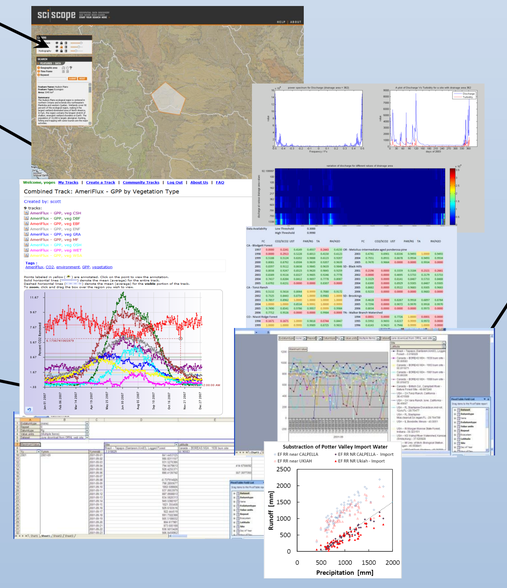
Models

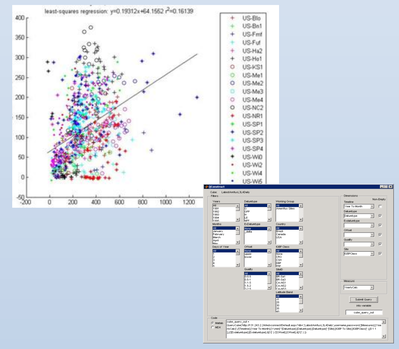
Archive Databases

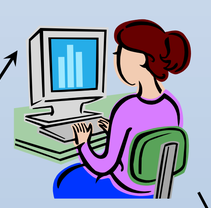


Data Cubes

Visualizations







Published scientific results

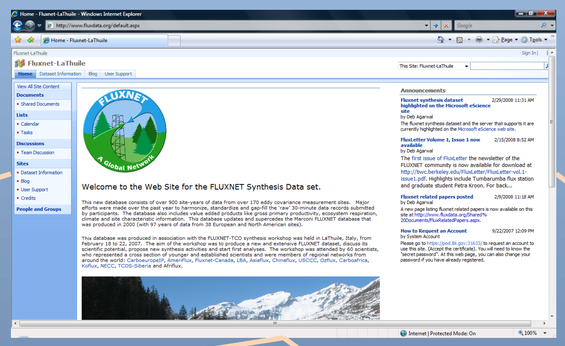
Fluxnet data server – www.fluxdata.org

- Common data repository capturing provenance information and common data processing
- Ancillary and biological data incorporated
- Data browsing, mining, and plotting capabilities
- Data repository that is easy to maintain, load with data, and expand capabilities
- Collaboration space to enable interaction between PIs and proposers and among proposal teams
- Data download and access (protected and tracked)

Reduce cut & paste manipulation of data

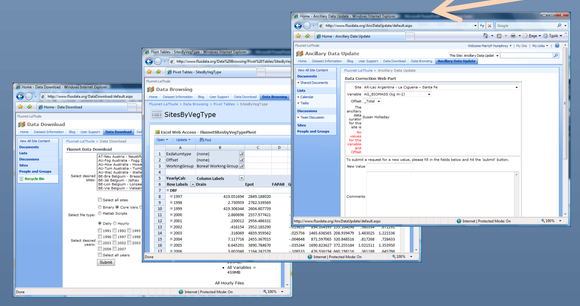
Dramatically increase the scale of the data that can be analyzed

Collaborative Interface with Customizable views/functionality based on role(s)

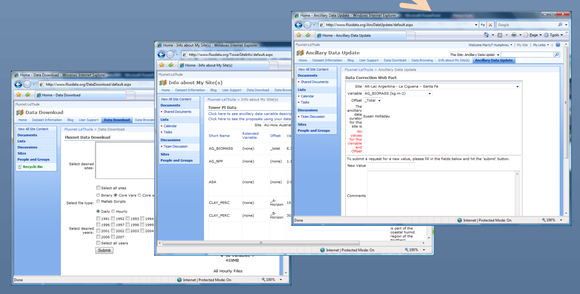


Powered by Microsoft Office Sharepoint Server 2007

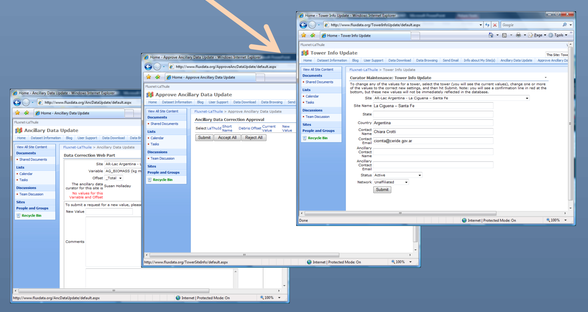
For the public: Dataset and tower information, user support , data summaries, list of synthesis teams, and published results



For the scientists analyzing the data: Download data, interactively explore data, provide ancillary data updates/corrections, and interact with tower owners



For the tower owners: Download curated tower data, review/update info about the tower, provide ancillary data updates/corrections, and interact with synthesis teams



For the data curators: Download data, provide ancillary data updates/corrections, reject/accept proposed data updates/corrections, and synchronize across data sources

Microsoft Research

http://www.fluxdata.org/

For further information e-mail: fluxdata-support@george.lbl.gov